9Ш

УДК 004.93’1

В.Ю. Шелепов, А.В. Жук
Институт проблем искусственного интеллекта МОН и НАН Украины, г. Донецк
shel@iai.dn.ua, juk@iai.dn.ua

Проблема распознавания длительностей как ритмических единиц музыкального произведения

В статье рассматривается задача распознавания длительностей как ритмических единиц музыкального произведения с точки зрения классической теории распознавания образов, выделяются основные сложности, которые появляются в процессе решения этой задачи, предлагаются возможные подходы к преодолению этих сложностей.

В настоящее время задача распознавания фонограмм музыкальных произведений (её также можно называть задачей распознавания звучащей музыки) стала актуальной благодаря развитию сети Интернет, мобильных технологий и робототехники. Задачу эту можно разбить на 3 подзадачи:
– сегментация звукового сигнала по признаку однородности частотного состава и распознавание выделенных сегментов;
– распознавание длительностей сегментов как ритмических единиц;
– определение общих ритмических характеристик музыкального произведения: размера и величины затакта.

Здесь следует отметить, что каждая из этих подзадач находит своё применение в разных и зачастую совершенно неожиданных областях. Так, в Интернете существует большое количество музыкальных поисковых сервисов, однако большинство из них осуществляют поиск музыкальных произведений по некоторой символьной информации, полученной от пользователя, и речь в данном случае идёт скорее о правильной организации и постоянном обновлении сетевых баз данных, связанных с музыкальными произведениями, чем о некотором процессе распознавания. В то же время, встречаются нетривиальные реализации перевода запроса пользователя из естественной для него формы (напев, простукивание ритма) в некоторую промежуточную форму. Наиболее продвинутым решением в данной области является проект MusicBrainz, использующий запатентованный алгоритм TRM фирмы Relatable, применяемый для организации акустических отпечатков. В сфере мобильных устройств также имеются похожие решения. Так, компания SonyEricsson предоставляет в своих устройствах функцию TrackID, которая позволяет по записанному фрагменту музыкального произведения осуществить его идентификацию в сетевой базе данных. Компания Motorola также заявляет в своих устройствах функцию SongID (аналог TrackID), однако обнаружить работающую реализацию данной технологии в присутствующих на рынке моделях телефонов этой компании не удалось. Здесь уместно отметить, что во всех рассмотренных случаях распознавание музыкального произведения в смысле перевода его в символьную форму представления (например, нотный

текст) не производится. Вместо этого выполняется выделение в произведении некоторых характерных признаков (например, ритмических характеристик) и дальнейший поиск производится уже по ним.

Поиск решения задачи определения ритмической структуры музыкального произведения активно ведётся в робототехнике, а результаты обкатываются на различных музицирующих и танцующих роботах, регулярно появляющихся на международных выставках в исполнении ведущих фирм, специализирующихся в области робототехники и электроники.

В то же время задача распознавания фонограмм музыкальных произведений в смысле перевода их в нотный текст на сегодняшний день не решена. Среди разработчиков наконец-то стало появляться понимание того, что по своей сложности эта задача приближается к задаче распознавания речи и является, безусловно, задачей искусственного интеллекта.

Некоторые варианты решений для задачи сегментации звукового сигнала по признаку однородности частотного состава приведены в [1-3].

Целью же данной работы является обзор возможных проблем, встающих при попытке решения задачи распознавания длительностей как ритмических единиц музыкального произведения и формализация этой задачи с позиций классической теории распознавания образов.

Эту задачу можно сформулировать так: для заданного набора временных продолжительностей звуков определить соответствующие длительности (в смысле ритмических единиц музыкального произведения) с учётом заданных априорно или определённых в процессе обработки общих ритмических характеристик музыкального произведения (темп, размер и затакт). Причём на данном этапе общие ритмические характеристики считаются постоянными величинами для рассматриваемого музыкального произведения.

Следует отметить, что подробное рассмотрение задачи автоматического определения общих ритмических характеристик музыкального произведения и методов её решения выходит за рамки данной статьи.

Постановка задачи

Очевидно, целесообразно попытаться формализовать рассматриваемую задачу с точки зрения классической теории распознавания образов. Это требует от нас указания признаков распознавания, количества распознаваемых системой классов, законов распределения признаков внутри каждого из классов (априори ясно, что признаки будут вероятностными).

Признак распознавания для данной задачи есть только один – это временная продолжительность звука, длительность которого необходимо определить. С одной стороны, наличие единственного признака распознавания избавляет разработчиков от необходимости оптимизации системы признаков. Однако известно [4], что эффективность системы распознавания увеличивается прямо пропорционально количеству используемых признаков. Очевидно, что эффективность рассматриваемой системы будет всецело зависеть от степени перекрытия классов в пространстве единственного имеющегося признака.

Количество различаемых классов в рамках данной задачи можно определить исходя из следующих соображений:

1) в большинстве музыкальных произведений используются длительности от целой до 64-й;

2) наряду со стандартными длительностями часто используются длительности увеличенной продолжительности – длительности с точкой и длительности с двойной точкой;

3) помимо вышеперечисленных длительностей также широко используются кортежи длительностей: триоли, пентоли, септоли и т.д.;

4) в некоторых ситуациях используются альтернативные кортежи – дуоли, квадроли, секстоли и т.д.

Если распознавать лишь указанные длительности, то количество классов оказывается равным

$$7 \cdot (1+2+3+3) = 63 .$$

Для того чтобы оценить степень перекрытия классов, необходимо, в первую очередь, представить эти классы в пространстве признаков. Между основными классами длительностей в музыке устанавливаются соотношения, показанные в табл. 1.

Таблица 1 – Соотношение между временными продолжительностями некоторых рассматриваемых длительностей

Название базового класса	Базовая длительность		Длительность с точкой		Один элемент кортежа триолей	
	t	$\log_2 t$	t	$\log_2 t$	t	$\log_2 t$
Целая	64	6	96	6.585	$\frac{128}{3}$	5,415
Половинка	32	5	48	5.585	$\frac{64}{3}$	4,415
Четверть	16	4	24	4.585	$\frac{32}{3}$	3,415
Восьмая	8	3	12	3.585	$\frac{16}{3}$	2,415
Шестнадцатая	4	2	6	2.585	$\frac{8}{3}$	1,415
Тридцать вторая	2	1	3	1.585	$\frac{4}{3}$	0,415
Шестьдесят четвёртая	1	0	$\frac{3}{2}$	0.585	$\frac{2}{3}$	– 0,585

Однако иметь дело с таким расположением классов в пространстве признаков неудобно. Гораздо лучше было бы, если бы основные классы в пространстве признаков были расположены более или менее равномерно. Этого эффекта для основных классов длительностей можно достичь, предварительно прологарифмировав по основанию 2 значения временных продолжительностей длительностей.

При этом для длительностей с точкой вместо положения точно посередине между центрами классов основных длительностей получился сдвиг, соответствующий

$$\log_2(1{,}5) = 0{,}585 .$$

Таким образом, на логарифмической шкале длительности с точкой оказываются незначительно смещёнными в сторону большей длительности.

И здесь появляются первые проблемы: к примеру, центры классов триольных длительностей на логарифмированной шкале оказываются очень близкими к длительностям с точкой. Та же картина наблюдается и в случаях с другими «кортежными» длительностями.

В связи с рассмотренной ситуацией расположения центров классов длительностей на логарифмической шкале признака распознавания, ключевую роль начинают играть законы распределения признака распознавания внутри введенных классов. На сегодняшний день объём проведенных исследований не позволяет установить экспериментальным путём эти законы, однако уже сейчас можно сказать, что степень перекрытия классов нарастает по мере добавления различных «дополнительных» классов длительностей. Так, если базовые классы практически не пересекаются, то при добавлении классов, соответствующих длительностям с точкой, пересечение становится уже достаточно значительным, чтобы порождать регулярные ошибки распознавания. При добавлении же классов, соответствующих «кортежным» длительностям, их становится очень трудно отличать от длительностей с точкой вследствие сильного перекрытия и близкого расположения их центров в пространстве признака распознавания.

Причин такого сильного перекрытия классов несколько, носящих как объективный, так и субъективных характер. В случае работы со звуковым файлом (что подразумевает его предварительную сегментацию по признаку однородности частотного состава) границы оказываются определёнными с некоторой ошибкой вне зависимости от применяемого метода сегментации. Поскольку эти ошибки зависят от величин используемого окна сегментации, фазового сдвига и реального периода основного тона сигнала, то величина их является случайной. Далее, априори ясно, что любые две «одинаковые» длительности, сыгранные музыкантом, при точном анализе будут отличаться друг от друга. Кроме того, в случае анализа записи живого исполнителя, временная продолжительность длительностей одного класса может изменяться в зависимости от эмоционального состояния исполнителя и семантической нагрузки, которую он вкладывает в воспроизведение. Проведённые исследования показали, что разброс внутриклассовых значений, порождаемый каждым из этих источников в отдельности, приводит к тому, что при полном наборе классов с достаточной точностью распознаются только основные классы длительностей.

Вполне возможно, что, если бы удалось определить закон распределения, с учетом каждого из возможных источников ошибок, удалось бы построить достаточно точную систему распознавания длительностей без обучения, однако возможен и другой подход к решению данной задачи.

Известно [4], что эффективность классификатора возрастает при уменьшении количества распознаваемых классов. Анализируя музыкальные произведения, можно заметить, что, как правило, в них используется только часть рассмотренных ранее классов. И в то же время, для каждого музыкального произведения этот набор классов будет отличаться. Получается, что в данной ситуации количество классов, которые должна распознавать система, не известно, хотя оно и значительно меньше, чем в рассмотренном выше случае системы без обучения. Такая ситуация характерна для систем распознавания с обучением без учителя. В то же время, характер ошибок меняется от фонограммы к фонограмме, что не позволяет использовать данные обучения на одной фонограмме для распознавания другой. Кроме того, следует отметить тот факт, что от фонограммы к фонограмме будет меняться положение центров классов в пространстве признаков, что связано, в основном, с отличиями в темпах.

Одним из наиболее привлекательных решений в данной ситуации представляется следующее: необходимо выполнить обучение классификатора без учителя, используя все длительности обрабатываемого музыкального произведения в качестве обучающей выборки. По результатам этого обучения можно будет определить положение центров классов длительностей, а затем – выполнить распознавание объектов этой же выборки. В результате обучения классификатора без учителя будет получен только набор кластеров, по которым будут распределены объекты обучаю-

щей выборки. Поэтому для определения положения центров классов в пространстве признаков системе необходимо будет предоставлять дополнительную априорную информацию о соответствии хотя бы одного из полученных кластеров реальному классу длительностей. Учитывая имеющиеся соотношения между классами длительностей, такой информации должно быть достаточно для восстановления соответствия в случае всех полученных в результате обучения кластеров.

Учитывая отсутствие информации о законах распределения признаков распознавания внутри каждого из классов длительностей, до проведения необходимого исследования можно предположить, что вероятность принадлежности к некоторому классу равна 1, если значение признака распознавания совпадает с центром класса, равномерно убывает с удалением от центра класса, и равна 0, если значение признака распознавания совпадает с центром соседнего класса, причём отсутствие сведений об априорной вероятности и платах за ошибки распознавания приводит к тому, что граница между классами будет располагаться как раз посередине между центрами соседних классов. График изменения вероятности отнесения к двум соседним классам представлен на рис. 1.

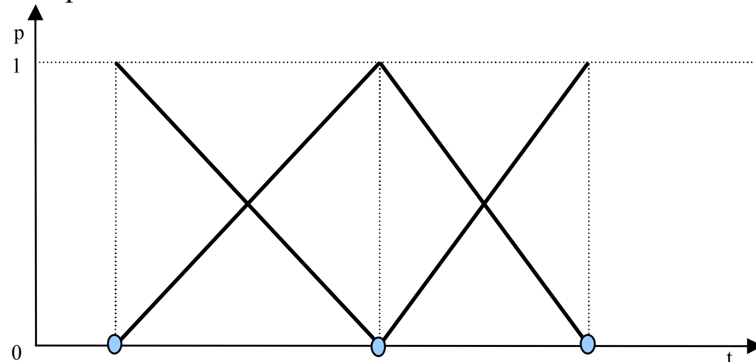


Рисунок 1 – Изменение вероятности отнесения объекта к двум соседним классам в соответствии с допущением, принятым в данной работе

Для удобства дальнейшей обработки параметр, характеризующий правильность принятого классификатором решения, вероятность ошибки распознавания в классической теории распознавания [5], представляется целесообразным модифицировать так, чтобы он отражал не только величину ошибки, но и её отклонение. Эту величину, конечно, нельзя будет называть вероятностью, поскольку её значение будет находиться в интервале [– 1,1]. В дальнейшем в тексте данной статьи описанная выше величина будет называться мерой ошибки распознавания и будет обозначаться как $\tilde{p}$.

Распознавание в соответствии с описанными выше законами распределения длительностей производится путём определения наиболее близкого к рассматриваемому образу центра класса. Результат распознавания целесообразнее всего выразить двойкой вида

$$R = (q, \tilde{p}), \tag{1}$$

где q – номер распознанного класса длительностей.

Как показали проведенные исследования, построенный подобным образом классификатор длительностей может ошибаться только в пределах соседних классов, что позволяет ввести альтернативное решение следующим образом:

$$\overline{R} = \begin{matrix} (q-1, 1-|\tilde{p}|),\ если\ \tilde{p} < 0 \\ (q+1, -1+\tilde{p}),\ если\ \tilde{p} > 0 \end{matrix}. \tag{2}$$

После распознавания зачастую оказывается необходимым выполнение коррекции полученных результатов. Данная коррекция будет заключаться в выборе для каждого объекта между полученным (1) и альтернативным (2) решением и может быть основана на следующих принципах:

1) минимизации количества синкоп в музыкальном произведении;

2) минимизации суммарной меры ошибки распознавания внутри каждого такта распознанной последовательности длительностей с учётом 1).

На основании изложенных в данной статье соображений авторами планируется разработка алгоритмов обучаемого классификатора для определения центров классов, классификатора без обучения для предварительного распознавания длительностей и апостериорного корректора результатов распознавания. Результаты работ по этому направлению будут приведены в последующих статьях по данной тематике.

Выводы

В статье рассмотрена задача распознавания длительностей как ритмических единиц музыкального произведения с точки зрения классической теории распознавания образов, выделены основные проблемы, которые появляются в процессе решения этой задачи, предложены возможные подходы к преодолению этих проблем. Основные сложности в работе с данной предметной областью, как показано в статье, связаны с наличием единственного признака распознавания и сильным перекрытием классов в случае наиболее общей постановки задачи. Для преодоления этих сложностей предлагается использовать двухпроходный классификатор, первый проход которого предназначен для определения центров классов для распознаваемого музыкального произведения, а второй – собственно для распознавания на основании полученных в результате первого прохода данных, причём обучающая выборка для первого и классифицируемая выборка для второго прохода совпадают. В дальнейшем на основании изложенных в статье соображений авторами планируется разработка алгоритмов автоматического распознавания длительностей и программных решений на их основе.

Литература

1. Жук А.В. Распознавание последовательности звуков одноголосной мелодии по высоте // Искусственный интеллект. – 2006. – № 2. – С. 305-312.
2. Жук А.В. Алгоритмы оценки частоты основного тона сигнала // Известия ТРТУ. Тематический выпуск «Интеллектуальные и многопроцессорные системы». – 2006. – № 16 (71). – С. 97-101.
3. Жук А.В. Алгоритм беспорогового определения частотного состава звукового сигнала // Искусственный интеллект. – 2007. – № 4. – С. 364-372.
4. Белозерский Л.А. Введение в системы автоматического распознавания. – Киев: Наук. думка, 2005. – 434 с.
5. Горелик А.Л., Скрипкин В.А. Методы распознавания. – М.: Высшая школа,1989. – 231 с.

В.Ю. Шелепов, О.В. Жук

Проблема розпізнавання тривалостей як ритмічних одиниць музичного твору

У статті розглядається завдання розпізнавання тривалостей як ритмічних одиниць музичного твору з позиції класичної теорії розпізнавання образів, виділяються основні складності, що з’являються в процесі вирішення цього завдання, пропонуються можливі підходи до розв’язання цих складностей.

V.Yu. Shelepov, A.V. Zhuk

The Duration Recognition Task in the Sense of «Rhythmical Units of Musical Composition»

The article is devoted to the duration (in the sense of «rhythmical units of musical composition») recognition task. This task is observed from the point of classical pattern-recognition theory. The main complications, which can appear during this task solution are marked, and possible approaches to their overcoming are proposed.

Статья поступила в редакцию 10.07.2008.